Exam 1 Review Solutions

1

Medical records show that, among patients suffering from disease D, 75% will die of it. For each situation described below: (i) define an appropriate random variable to model the situation; (ii) give the values that the random variable can take on; (iii) find the probability that the random variable is greater than 3; (iv) state any assumptions you need to make. Be sure to label parts (i) —(iv) clearly for full credit!

(a)

Out of 10 people suffering from D, x people will survive.

Solution:

(i) This would be a binomial r.v. with $n = 10$ people and $p = 1 - 0.75 = 0.25$

$$X \sim Bin(10, 0.25)$$

(ii) All ten can die, all ten can survive, or any in between.

$$X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

(iii)

$$P(X > x) = 1 - P(X \leq x) = 1 - \sum_{n=0}^{x} \binom{n}{k} * p^k * (1-p)^{n-k}$$

$$P(X > 3) = 1 - \sum_{k=0}^{3} [\binom{10}{k} * (0.25)^k * (0.75)^{10-k}]$$

$$= 0.2241$$

(iv) Assume trials (people) are independent of one another and each trial has the same probability.

(b)

People die from D at a rate of 4 people per day.

Solution:

(i) This is a Poisson r.v. with a rate $\lambda = 4$

$$X \sim Pois(4)$$

(ii) $X \in \{0, 1, \dots\}$

(iii)

$$P(X > 3) = \sum_{i=4}^{\infty} \frac{\lambda^i * e^{-\lambda}}{i!} = 1 - \sum_{i=0}^{3} \frac{\lambda^i * e^{-\lambda}}{i!} = 1 - \sum_{i=0}^{3} \frac{4^i * e^{-4}}{i!}$$

$$= 0.567$$

(iv)(Note that we didn't cover these explicitly in class, so I won't test you on them):

1. The probability that one person dies within a small time interval is approximately proportional to the size of that interval.
2. The probability of two deaths occurring in the same narrow interval is negligible.
3. The probability of a death within a certain interval does not change over different intervals.
4. The probability of a death in one interval is independent of the probability of a death in any other non-overlapping interval.

(c)

The average amount of time, t (in hours), until the first person dies on a given day is 3.

Solution :

This is an exponential distribution problem. It will not be covered on this exam...

(i) This is an exponential distribution with mean of 3 so $\lambda = 1/3$

$$X \sim Exp\left(\frac{1}{3}\right)$$

(ii) $X \in [0, \infty)$

(iii)

$$P(X > 3) = \int_3^{\infty} \lambda e^{-\lambda x} dx = 1 - \int_0^3 \lambda e^{-\lambda x} dx = 1 - \int_0^3 \frac{1}{3} e^{-x/3} dx$$

$$= 1 - \frac{1}{3}\left[\frac{e^{(-x/3)}}{\frac{-1}{3}}\right]_0^3 = 1 + \left[e^{(-x/3)}\right]_0^3$$

$$= 1 + [e^{-1} - 1] = e^{-1} = 0.368$$

(iv) The exponential distribution describes the amount of time it takes for a death to occur. We assume that deaths are indpendent or one another.

2

(a)

Let $f(x)$ be the probability distribution function for a continuous random variable X. Then f(x) > 0 for all x.

Solution:

FALSE.

$f(x) \geq 0$ for all x.

(b)

For any set B, $P(B \mid B^C) = 0$.

Solution:

FALSE.

$P(B \mid B^C)$ is *undefined* when $P(B^c) = 0$. When $P(B^c) \neq 0$

$$P(B \mid B^C) = \frac{P(B \cap B^C)}{P(B^C)} = 0,$$

because $P(B \cap B^C) = 0$.

(c)

(c) See PDF For question.

Solution:

TRUE.

The line $x[\ ,2]$ extracts the second column of x.

(d)

If $P(A) = 0.5$ and $P(B) = 0.5$, then $P(A \cap B) \leq 0.25$.

Solution:

FALSE.

Let $P(A \mid B) = 0.9$. Then $P(A \cap B) = P(B)P(A \mid B) = (0.5)(0.9) = 0.45 > 0.25$.

(e)

The median for the following dataset is 6: 3, 4, 12, 13, 4, 10, 6, 15, 1, 2, 2, 12.

Solution:

List the set in order: $\{1, 2, 2, 3, 4, 4, 6, 10, 12, 12, 13, 15\}$

FALSE.

The median of the set is $\frac{4+6}{2} = 5$.

(f)

Suppose a sample is taken, and the variable of interest z is measured for each unit in the sample. If the distribution of z is bimodal as shown (see PDF), the sample median would be the best measure of center.

Solution:

FALSE.

It would be most reasonable to use the two modes centers of measure.

(g)

For *any* data set $x_1, x_2, \ldots, x_n$, the variance can be written as $\frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - n\bar{x}^2)$ (Prove or disprove).

Solution:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \bar{x}^2 \sum_{i=1}^{n} 1\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x} * (\bar{x}n) + \bar{x}^2 * n\right)$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i^2 - n\bar{x}^2)$$

TRUE.

(h)

For sets A and B, $P(A \mid B) = P(B \mid A)$.

Solution:

FALSE.

Consider $P(A), P(B) > 0$ with $P(A) \neq P(B)$

Then,

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)} \neq P(B \mid A)$$

3

The axioms of probability state that, for any event E, $P(E) \geq 0$, but they do not explicitly state that $P(E) \leq 1$. Prove the claim that, for any event E, $P(E) \leq 1$. Why do you think that this claim is omitted from the axioms?

Solution:

We know that $E \subset \Omega$, that $P(\Omega) = 1$, and that, in general, if $A \subset B$ then $P(A) \leq P(B)$ *. Thus, $P(E) \leq P(\Omega) = 1$.

*Can you prove this??

4

A box contains 5 defective bulbs, 10 partially defective bulbs (the bulb starts to fail after 10 hours of use) and 25 perfect bulbs. If a bulb is selected at random from the box, tested, and it does not fail immediately, what is the probability that it is perfect?

Solution:

Consider the events of picking a bulb at random from the 40 bulbs,
$D =$ Defective, $PD =$ Partially Defective, $P =$ Perfect

We are given, $$P(D) = \frac{5}{40}$

$$P(PD) = \frac{10}{40}$$

$$P(P) = \frac{25}{40}$$

We are asked to find the probability that a bulb is perfect given that it is not defective. That is, $P(P \mid D^C)$.

$$P(P \mid D^C) = \frac{P(P \cap D^C)}{P(D^C)}$$

Note that $P(D^C) = 1 - P(D) = 1 - \frac{5}{40} = \frac{35}{40}$ and $P(P \cap D^C) = \frac{25}{40}$.

$$P(P \mid D^C) = \frac{\frac{25}{40}}{\frac{35}{40}} = \frac{25}{35}$$

5

Fix a set B, such that $P(B) > 0$. Show that the definition of conditional proba- bility $(P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ satisfies the axioms of probability theory.

Solution:

1. We need to show that $P(\Omega \mid B) = 1$. Note that
$$P(\Omega \mid B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$
2. Here, we need to show that $P(E \mid B) \geq 0$ for any event E. But note that
$$P(E \mid B) = \frac{P(E \cap B)}{P(B)} \geq 0$$
because both the numerator and denominator are greater than zero.

3. We need to show that, for any E_1 and E_2 such that $E_1 \cap E_2 = \emptyset$,
   $P(E_1 \cup E_2 \mid B) = P(E_1 \mid B) + P(E_2 \mid B)$.

$$P(E_1 \cup E_2 \mid B) = \frac{P([E_1 \cup E_2] \cap B)}{P(B)} = \frac{P([E_1 \cap B] \cup [E_2 \cap B])}{P(B)} = \frac{P(E_1 \cap B) + P(E_2 \cap B)}{P(B)}$$

6

Suppose that in the field of exercise science, 1 out of 100 claims are true (let C = 'a given claim is true'. An example of a claim in exercise science: "jogging 3 times a week for 20 minutes increases cardiovascular health."). In practice, the only way we know whether a claim is true is if we investigate/research the claim. Claims that are investigated and thought to be true are published as a "research finding". Of course, research findings can be wrong. Given that a research claim in exercise science is true, a research finding correctly reports it 50% of the time (let R = 'research finding says the claim is true'). Given that a research claim in exercise science is false, a research finding incorrectly reports it 40% of the time.

(a)

What is the probability that a research finding says that a research claim in exercise science is true?

Solution:

We are given,

$P(C) = \frac{1}{100}$, $P(C^C) = \frac{99}{100}$, $P(R \mid C) = 0.5$, $P(R \mid C^C) = 0.4$

C and C^C span sample space so use the Law of Total Probability,

$$P(R) = P(R \mid C) * P(C) + P(R \mid C^C) * P(C^C)$$

$$P(R) = (0.5) * (0.01) + (0.4) * (0.99) = 0.401$$

$$P(R) = 0.401$$

(b)

What is the probability that a research claim is true given that a research finding says that it is?

Solution:

We are asked to find $P(C \mid R)$

Use Bayes' Theorem,

$$P(C \mid R) = \frac{P(R \mid C) * P(C)}{P(R)}$$

$$= \frac{(0.5) * (0.01)}{(0.401)}$$

$$P(C \mid R) = 0.0125$$

(c)

Interpret your answer for (b)? Is it alarming? Why? (Note that this scenario is not implausible!)

Solution:

Yes, it is alarming that there is just over a 1% chance that a research claim is true given that a research finding says that it is. However, this is because a combination of the fact that 99% of the overall claims are false and only 50% of the 1% of true claims are actually deemed so by research findings.

7

Let X = the leading digit of a randomly selected number from a large accounting ledger. So, for example, if we randomly draw the number 20,695, then X = 2. People who make up numbers to commit accounting fraud tend to give X a (discrete) uniform distribution, i.e., $P(X = x) = 1/9$, for $x \in 1, \ldots, 9$. However, there is empirical evidence that suggests that "naturally occurring" numbers (e.g., numbers in a non-fraudulent accounting ledger) have leading digits that do not follow a uniform distribution. Instead, they follow a distribution defined by:

$$f(x) = \log_{10} \frac{x+1}{x}, \quad x = 1, 2, \ldots, 9.$$

Show that $f(x) = P(X = x)$ is, in fact, a probability distribution function.

Solution:

To show $f(x)$ is pdf, show that

$$P(X = x) \geq 0 \ , \ x = \{1, 2, \ldots, 9\}$$

The argument inside the logarithm for all x is greater than 1 so
$P(X = x) \geq 0 \ for\, x = \{1, 2, \ldots, 9\}$

and now show

$$\sum_{x=1}^{9} f(x) = \sum_{x=1}^{9} \log_{10}\left(\frac{x+1}{x}\right) = 1$$

$$\sum_{x=1}^{9} \log_{10}\frac{x+1}{x} = \log_{10}\frac{2}{1} + \log_{10}\frac{3}{2} + \log_{10}\frac{4}{3} + \log_{10}\frac{5}{4} + \log_{10}\frac{6}{5} + \log_{10}\frac{7}{6} + \log_{10}\frac{8}{7}$$

$$= \log_{10}\left(\frac{2*3*4*5*6*7*8*9*10}{1*2*3*4*5*6*7*8*9}\right) = \log_{10} 10 = 1$$

8

Consider the following R code:

```
In [11]: x = matrix(rgeom(16,0.6), ncol = 4)
         x
```

2	0	2	5
0	0	0	0
0	0	1	1
1	3	1	0

Interpret the value in $x[4, 2]$

Solutions:

$x[4, 2] = 3$